

# Binary cumulant varieties

Bernd Sturmfels and Piotr Zwiernik

**Abstract.** Algebraic statistics for binary random variables is concerned with highly structured algebraic varieties in the space of  $2 \times 2 \times \cdots \times 2$ -tensors. We demonstrate the advantages of representing such varieties in the coordinate system of binary cumulants. Our primary focus lies on hidden subset models. Parametrizations and implicit equations in cumulants are derived for hyperdeterminants, for secant and tangential varieties of Segre varieties, and for certain context-specific independence models. Extending work of Rota and collaborators, we explore the polynomial inequalities satisfied by cumulants.

**Mathematics Subject Classification (2010).** Primary: 13P25; Secondary: 05A40, 14Q15, 60C05.

**Keywords.** Algebraic statistics, cumulants, moments, binary data, context-specific independence, hyperdeterminant, Segre variety, secant variety.

## 1. Introduction

Cumulants have a long and interesting history dating back to Thorvald N. Thiele, a Danish mathematician, who introduced them in 1889. See [8] for a historical perspective. The main motivation to study them was that multivariate probability distributions are often easier to analyze when expressed in terms of cumulants. Moreover, cumulants are especially useful when dealing with the normal distribution, and hence they are a critical tool in asymptotic statistics (see e.g. [2, 11, 23, 26]). Various invariance properties of cumulants make them interesting also from an algebraic or combinatorial point of view. Rota and his collaborators [1, 21] developed a combinatorial theory of cumulants, and, more recently, Pistone and Wynn introduced *cumulant varieties* [16] into algebraic statistics. These concepts gave rise to umbral calculus [20], an approach to combinatorial sequences using cumulants.

Building on this circle of ideas, we show how cumulants can be used to study algebraic varieties in tensor spaces. Thus, cumulants can be also used outside of the probabilistic context where we deal with sequences of nonnegative numbers summing to 1. Here we focus on binary states. Let

---

This work was conducted during the Spring 2011 program “*Algebraic Geometry with a View Towards Applications*” at the Institute Mittag-Leffler in Djursholm, Sweden. The first author was partially supported by KTH Stockholm and the US National Science Foundation (DMS-0968882). The second author gratefully acknowledges support from the AXA Mittag-Leffler Fellowship Project, sponsored by the AXA Research Fund.

$P = [p_I]_{I \subseteq [n]}$  be an  $n$ -dimensional  $2 \times 2 \times \cdots \times 2$  table of complex numbers such that  $\sum p_I = 1$ . We call such tensors *distributions*. In statistical contexts one assumes in addition that the  $p_I$  are real and nonnegative in which case we call them *probability distributions*. In algebraic statistics, the probabilities  $p_I$  form the coordinates of the ambient space containing statistical models. For an introduction to this geometric point of view see [3].

We represent the distribution  $P$  by the *probability generating function*

$$P(x) = \sum_{I \subseteq [n]} p_I \prod_{i \in I} x_i.$$

Here  $[n] = \{1, 2, \dots, n\}$  and we identify our tables with functions on subsets of  $[n]$ . In the probabilistic context we occasionally refer to the random vector  $X = (X_1, \dots, X_n)$  with values in  $\{0, 1\}^n$  and distribution  $P$ . We use here the natural identification of a subset  $I \subseteq [n]$  with its support vector. An alternative representation of  $P$  is the table of moments  $M = [\mu_I]_{I \subseteq [n]}$ , where

$$\mu_I = \sum_{J \supseteq I} p_J. \quad (1)$$

The *moment generating function* is a square-free polynomial in  $n$  unknowns:

$$M(x) = P(x_1 + 1, \dots, x_n + 1) = \sum_{I \subseteq [n]} \mu_I \prod_{i \in I} x_i. \quad (2)$$

The logarithm of the moment generating function gives the cumulants:

$$K(x) = \sum_{I \subseteq [n]} k_I \prod_{i \in I} x_i := \log(M(x)). \quad (3)$$

Note that  $\mu_\emptyset = 1$  and  $k_\emptyset = 0$ . For the logarithm we use the familiar series  $\log(1+t) = \sum_{i=1}^{\infty} (-1)^{i-1} t^i / i$ . That expansion is understood modulo the ideal  $\langle x_1^2, x_2^2, \dots, x_n^2 \rangle$ . The moments can then be recovered from the cumulants via

$$M(x) = \exp(K(x)). \quad (4)$$

The transformations (3) and (4) between moments  $\mu_I$  and cumulants  $k_I$  can be written as explicit combinatorial formulas (see e.g. [11, §2.3], [21, 23]). Given any  $I \subseteq [n]$ , let  $\Pi(I)$  be the lattice of all set partitions of  $I$ . We have

$$k_I = \sum_{\pi \in \Pi(I)} (-1)^{|\pi|-1} (|\pi| - 1)! \prod_{B \in \pi} \mu_B. \quad (5)$$

The sum is over partitions of  $I$ , the product is over blocks of a partition, and  $|\pi|$  denotes the number of blocks of  $\pi$ . The moments in terms of cumulants are

$$\mu_I = \sum_{\pi \in \Pi(I)} \prod_{B \in \pi} k_B \quad \text{for all } I \subseteq [n]. \quad (6)$$

For instance,  $I = \{1, 2, 3\}$  has five partitions 123, 1|23, 2|13, 12|3, and

$$\begin{aligned} k_{123} &= \mu_{123} - \mu_1 \mu_{23} - \mu_2 \mu_{13} - \mu_{12} \mu_3 + 2\mu_1 \mu_2 \mu_3, \\ \mu_{123} &= k_{123} + k_{12} k_3 + k_{13} k_2 + k_{23} k_1 + k_1 k_2 k_3. \end{aligned} \quad (7)$$

The transformation from (6) to (5) is the *Möbius inversion* on the *partition lattice*  $\Pi([n])$ , as seen in enumerative combinatorics [24, Exercise 3.44].

This article is organized as follows. In Section 2 we study the expression of *hyperdeterminants* in terms of cumulants. In Section 3 we show that  $SL(2)^n$ -invariant tensor varieties are defined by  $\mathbb{Z}^n$ -homogeneous polynomials in the higher order cumulants  $k_I$  with  $|I| \geq 2$ . Section 4 concerns secants and tangents of the Segre variety, and we show (in Theorem 4.1) that the *tangential variety* becomes toric in cumulant coordinates. A conceptual explanation for this arises from our theory of *hidden subset models*, developed in Section 5. Here the main result is Theorem 5.1. Section 6 offers an algebraic study of the *context-specific independence models* due to Georgi and Schliep [6]. Section 7 explores the *semialgebraic constraints* on cumulants arising from probabilities, and it addresses a conjecture proposed in [1].

## 2. Hyperdeterminants

One of the most intriguing polynomial functions on  $2 \times 2 \times \cdots \times 2$ -tables is the *hyperdeterminant*  $\text{Det}(P)$ , which is a generalization of the determinant of a  $2 \times 2$  matrix. The hyperdeterminant, first introduced by Cayley in 1843, has many equivalent definitions (see [5]). One of them states that  $\text{Det}(P)$  is the (unique up to scaling) irreducible polynomial in the  $p_I$  that vanishes whenever the complex hypersurface defined by the equation  $P(x) = 0$  has a singular point in  $\mathbb{C}^n$ . Algebraically, the hyperdeterminant  $\text{Det}(P)$  is obtained by eliminating the  $n$  unknowns  $x_1, x_2, \dots, x_n$  from the  $n + 1$  equations

$$P(x) = \frac{\partial P}{\partial x_1}(x) = \frac{\partial P}{\partial x_2}(x) = \cdots = \frac{\partial P}{\partial x_n}(x) = 0.$$

According to [5, §14.2],  $\text{Det}(P)$  is a homogeneous polynomial of degree  $C_n$  in the  $2^n$  unknowns, where  $\sum_{n=0}^{\infty} C_n z^n / n! = e^{-2x} / (1 - x)^2$ . So, the degrees of our hyperdeterminants are  $C_2 = 2, C_3 = 4, C_4 = 24, C_5 = 128$  etc.

We work in the  $(2^n - 1)$ -dimensional affine space of distributions defined by  $\sum_I p_I = 1$ , or  $\mu_{\emptyset} = 1$ . We seek to express the hyperdeterminant on that affine space in terms of the cumulants  $k_I$ . From such an expression one recovers a formula for  $\text{Det}(P)$  in terms of the original coordinates  $p_I$ , up to scaling, by using (1) and (5).

If  $n = 2$  then the hyperdeterminant is the determinant of a  $2 \times 2$ -matrix,

$$P = \begin{bmatrix} p_{\emptyset} & p_2 \\ p_1 & p_{12} \end{bmatrix}.$$

In statistics, this represents the *independence model* for two binary random variables, and we recover the well-known fact that independence is equivalent to vanishing of the covariance

$$\text{Det}(P) = p_{12}p_{\emptyset} - p_1p_2 = \mu_{12} - \mu_1\mu_2 = k_{12}.$$

The statistical meaning of larger hyperdeterminants will be discussed later. See, in particular, the context-specific independence model in Example 6.2.

If  $n = 3$  then, by [5, Proposition 14.1.7], the hyperdeterminant equals

$$\begin{aligned} \text{Det}(P) = & \mu_1^2 \mu_{23}^2 + \mu_2^2 \mu_{13}^2 + \mu_3^2 \mu_{12}^2 + \mu_{123}^2 + 4(\mu_1 \mu_2 \mu_3 \mu_{123} + \mu_{12} \mu_{13} \mu_{23}) \\ & - 2(\mu_1 \mu_2 \mu_{13} \mu_{23} + \mu_1 \mu_3 \mu_{12} \mu_{23} + \mu_2 \mu_3 \mu_{12} \mu_{13} + \mu_1 \mu_{23} \mu_{123} + \mu_2 \mu_{13} \mu_{123} + \mu_3 \mu_{12} \mu_{123}) \end{aligned}$$

Here we can use either  $\mu_I$  or  $p_I$  since  $\text{Det}(P)$  is  $\text{SL}(2)^3$ -invariant. The formula simplifies considerably after we replace moments by cumulants via (4) or (6):

$$\text{Det}(P) = k_{123}^2 + 4k_{12}k_{13}k_{23}. \quad (8)$$

This  $2 \times 2 \times 2$ -hyperdeterminant is also known as the *tangle*, and it appears in phylogenetics [25], quantum computation [12] and string theory [4].

The next case  $n = 4$  is much more challenging. According to Huggins *et al.* [10], the  $2 \times 2 \times 2 \times 2$ -hyperdeterminant has precisely 2,894,276 terms, when written as a polynomial of degree 24 in either probabilities  $p_I$  or moments  $\mu_I$ . However, the expansion of  $\text{Det}(P)$  in terms of cumulants  $k_I$  is much smaller. The following theorem is our main result in this section.

**Theorem 2.1.** *The  $2 \times \cdots \times 2$ -hyperdeterminant  $\text{Det}(P)$  is a polynomial function in the  $2^n - n - 1$  higher cumulants  $\{k_I : |I| \geq 2\}$ . It is homogeneous of degree  $\frac{1}{2}(C_n, C_n, \dots, C_n)$  in the  $\mathbb{Z}^n$ -grading given by  $\deg(k_I) = \sum_{i \in I} e_i$ , where  $e_i$  is the  $i$ -th unit vector of  $\mathbb{Z}^n$ . For  $n = 4$ , the hyperdeterminant  $\text{Det}(P)$  has precisely 13,819 monomials in the 11 unknowns  $k_I$ , all  $\mathbb{Z}^4$ -homogeneous of degree  $(12, 12, 12, 12)$ , and their total degrees range from 24 to 15.*

*Proof.* The expression of the hyperdeterminant in terms of the moments  $\mu_I$  coincides with the  $\mathcal{A}$ -discriminant (cf. [5]) of the moment generating function

$$M(x) = \sum_{I \subseteq [n]} \mu_I \prod_{i \in I} x_i = \exp(K(x)). \quad (9)$$

Here  $\mathcal{A}$  is the  $(n+1) \times 2^n$  matrix whose columns are the homogeneous coordinates of the vertices of the standard  $n$ -cube. Standard results on  $\mathcal{A}$ -discriminants ensure that  $\text{Det}(P)$  is homogeneous in the  $\mathbb{Z}^{n+1}$ -grading specified by  $\mathcal{A}$ , so, in particular, it is homogeneous in the coarser  $\mathbb{Z}^n$ -grading given by  $\deg(\mu_I) = \sum_{i \in I} e_i$ . Since the degree of  $\text{Det}(P)$  in the standard  $\mathbb{Z}$ -grading  $\deg(\mu_I) = 1$  equals  $C_n$ , as discussed above, we find that  $\text{Det}(P)$  is  $\mathbb{Z}^n$ -homogeneous of degree  $\frac{1}{2}(C_n, C_n, \dots, C_n)$ .

The map (6) from moments to cumulants respects the  $\mathbb{Z}^n$ -grading, and we conclude that the expansion of  $\text{Det}(P)$  in cumulants is  $\mathbb{Z}^n$ -homogeneous of the same degree  $\frac{1}{2}(C_n, C_n, \dots, C_n)$ . The first assertion that  $\text{Det}(P)$  does not depend on the first order moments  $k_1, \dots, k_n$  follows from Theorem 3.2.

We now come to the specific case  $n = 4$ . Here the proof was carried out by a computer calculation. We first set  $k_1, k_2, k_3$  and  $k_4$  to zero in the right hand side of (9) since  $\text{Det}(P)$  does not depend on these first-order cumulants. Our task is then to evaluate the  $\mathcal{A}$ -discriminant of the multilinear polynomial

$$\begin{aligned} M(x)|_{k_1=k_2=k_3=k_4=0} = & (k_{1234} + k_{12}k_{34} + k_{13}k_{24} + k_{14}k_{23})x_1x_2x_3x_4 \\ & + k_{123}x_1x_2x_3 + k_{124}x_1x_2x_4 + k_{134}x_1x_3x_4 + k_{234}x_2x_3x_4 \\ & + k_{12}x_1x_2 + k_{13}x_1x_3 + k_{14}x_1x_4 + k_{23}x_2x_3 + k_{24}x_2x_4 + k_{34}x_3x_4 + 1. \end{aligned}$$

This computation is done using Schläfli's formula [10, Prop. 3]. We obtained  $256k_{12}^6k_{13}^5k_{14}k_{23}k_{24}^5k_{34}^6 - 1024k_{12}^6k_{13}^4k_{14}^2k_{23}^2k_{24}^4k_{34}^6 + 1536k_{12}^6k_{13}^3k_{14}^3k_{23}^3k_{24}^4k_{34}^6 + \dots \text{many terms} \dots - k_{34}k_{12}^3k_{13}^3k_{14}^2k_{23}^2k_{24}^4k_{1234} + k_{12}^3k_{13}^3k_{14}^3k_{23}^3k_{24}^4k_{1234}.$

This expansion of  $\text{Det}(P)$  has 13819 terms, all of  $\mathbb{Z}^4$ -degree  $(12, 12, 12, 12)$ . The leading terms have total degree 24. The last terms have total degree 15.  $\square$

Ideals generated by hyperdeterminants arise in various applications. We advocate writing these in terms of cumulants. One such application, studied by Holtz-Sturmfels [9] and Oeding [14], concerns the relations among principal minors of a general symmetric  $n \times n$ -matrix  $A$ . If we write  $\mu_I$  for the minor with row and column indices  $I \subseteq [n]$ , and we treat the sequence  $[\mu_I]$  as a sequence of formal moments, then the corresponding moment generating function takes the special form

$$M(x) = \det(I + AX) \quad \text{where } X = \text{diag}(x_1, \dots, x_n).$$

Oeding [14] shows that the variety of such tables  $M = [\mu_I]$  is cut out by polynomials of degree 4. These polynomials are obtained by acting with the group  $SL(2)^n$  on the  $2 \times 2 \times 2$ -hyperdeterminants of all subtables. We reparametrize our variety of principal minors using the cumulant generating function:

$$K(x) = \log \det(I + AX) = \text{trace} \log(I + AX) = \text{trace} \left( \sum_{k=1}^n \frac{(-1)^{k+1}}{k} (AX)^k \right).$$

The coefficients  $k_I$  of the squarefree terms are sums over all cycle monomials in  $A$  that are supported on  $I$ . Their algebraic relations can be computed more easily than those among the principal minors. We demonstrate this for  $n = 4$ :

$$\begin{aligned} K(x) &= \sum_{I \subseteq [4]} k_I \prod_{i \in I} x_i \\ &= \sum_{i=1}^4 a_{ii} x_i - \sum_{i < j} a_{ij}^2 x_i x_j + 2 \sum_{i < j < k} a_{ij} a_{ik} a_{jk} x_i x_j x_k \\ &\quad - 2(a_{12} a_{13} a_{24} a_{34} + a_{12} a_{14} a_{23} a_{34} + a_{13} a_{14} a_{23} a_{24}) x_1 x_2 x_3 x_4 \end{aligned}$$

The prime ideal of algebraic relations among the coefficients is found to be

$$\begin{aligned} &\langle 4k_{12}k_{13}k_{23} + k_{12}^2k_{23}, 4k_{12}k_{14}k_{24} + k_{12}^2k_{24}, 4k_{13}k_{14}k_{34} + k_{13}^2k_{34}, 4k_{23}k_{24}k_{34} + k_{23}^2k_{34}, \\ &\quad 4k_{12}k_{13}k_{14}k_{234} + k_{123}k_{124}k_{134}, 4k_{12}k_{23}k_{24}k_{134} + k_{123}k_{124}k_{234}, \\ &\quad 4k_{13}k_{23}k_{34}k_{124} + k_{123}k_{134}k_{234}, 4k_{14}k_{24}k_{34}k_{123} + k_{124}k_{134}k_{234}, \\ &\quad 2k_{12}k_{13}k_{234} + 2k_{12}k_{23}k_{134} + 2k_{13}k_{23}k_{124} + k_{123}k_{1234}, \\ &\quad 2k_{12}k_{14}k_{234} + 2k_{12}k_{24}k_{134} + 2k_{14}k_{24}k_{123} + k_{124}k_{1234}, \\ &\quad 2k_{13}k_{14}k_{234} + 2k_{13}k_{34}k_{124} + 2k_{14}k_{34}k_{123} + k_{134}k_{1234}, \\ &\quad 2k_{23}k_{24}k_{134} + 2k_{23}k_{34}k_{124} + 2k_{24}k_{34}k_{123} + k_{234}k_{1234}, \\ &\quad -2k_{12}k_{13}k_{14}k_{1234} + k_{12}k_{13}k_{124}k_{134} + k_{12}k_{14}k_{123}k_{134} + k_{13}k_{14}k_{123}k_{124}, \\ &\quad -2k_{12}k_{23}k_{24}k_{1234} + k_{12}k_{23}k_{124}k_{234} + k_{12}k_{24}k_{123}k_{234} + k_{23}k_{24}k_{123}k_{124}, \\ &\quad -2k_{13}k_{23}k_{34}k_{1234} + k_{13}k_{23}k_{134}k_{234} + k_{13}k_{34}k_{123}k_{234} + k_{23}k_{34}k_{123}k_{134}, \\ &\quad -2k_{14}k_{24}k_{34}k_{1234} + k_{14}k_{24}k_{134}k_{234} + k_{14}k_{34}k_{124}k_{234} + k_{24}k_{34}k_{124}k_{134}, \\ &\quad k_{14}k_{123}k_{234} - k_{23}k_{124}k_{134}, k_{13}k_{124}k_{234} - k_{24}k_{123}k_{134}, k_{12}k_{134}k_{234} - k_{34}k_{123}k_{124}, \\ &\quad 4(k_{12}k_{13}k_{24}k_{34} + k_{12}k_{14}k_{23}k_{34} + k_{13}k_{14}k_{23}k_{24}) \\ &\quad - 2(k_{14}k_{123}k_{234} + k_{24}k_{123}k_{134} + k_{34}k_{123}k_{124}) - k_{1234}^2 \rangle \end{aligned}$$

These twenty polynomial correspond to the hyperdeterminantal relations in [9, Thm. 8]. They furnish a compact encoding of this codimension 5 variety.

### 3. Invariance and Independence

The algebraic relations in Section 2 did not involve any of the order one cumulants  $k_1, \dots, k_n$  and they were homogeneous with respect to the  $\mathbb{Z}^n$ -grading given by  $\deg(k_I) = \sum_{i \in I} e_i$ . In this section we argue that these properties hold for all statistically meaningful varieties in the space of  $2 \times \dots \times 2$ -tables.

To compute moments in (1) we used the convention that the (formal) random vector  $X = (X_1, \dots, X_n)$  has values in  $\{0, 1\}^n$ . Other authors prefer the choice  $\{-1, 1\}^n$ , and this leads to rather different formulas for the moments (see [1, Equation (2.1)]). A meaningful statistical model will not depend on such choices. Hence we are only interested in cumulant varieties that do not depend on such choices.

Suppose we replace each of our random variable  $X_i$  by a new variable  $X'_i$  which takes values  $a_i$  and  $b_i$  instead of 1 and 0. If the probability distribution is the same on both state spaces, then the cumulants are transformed via

$$k'_I = k_I \cdot \prod_{i \in I} (a_i - b_i) \quad \text{for all } I \subseteq [n] \text{ and } |I| \geq 2 \quad (10)$$

and  $k'_i = (a_i - b_i)k_i + b_i$  for  $i = 1, \dots, n$ . This result is purely algebraic and the above remains true if we replace probability distributions with any complex distributions. In geometric language, changing the values of the binary variables  $X_i$  corresponds to a natural action of the  $n$ -dimensional torus  $(\mathbb{C}^*)^n$  with coordinates  $a_i - b_i$  on the space  $\mathbb{C}^{2^n - n - 1}$  whose coordinates are the higher cumulants  $k_I$ ,  $|I| \geq 2$ . This action is compatible with the  $\mathbb{Z}^n$ -grading:

**Theorem 3.1.** *A subvariety of  $\mathbb{C}^{2^n - 1}$  is invariant under changing values of the  $X_i$  if and only if it is defined by  $\mathbb{Z}^n$ -homogeneous polynomials in  $k_I$  with  $|I| \geq 2$ .*

*Proof.* Let  $V$  be a subvariety in the space  $\mathbb{C}^{2^n - 1}$  whose coordinates are all the cumulants. Suppose that  $V$  is invariant under replacing the values  $(0, 1)$  of  $X_i$  by any  $(b_i, a_i)$ . If the new values satisfy  $a_i = b_i + 1$  then the higher cumulants  $k_I$ ,  $|I| \geq 2$ , remain unchanged but the vector  $(k_1, \dots, k_n)$  is shifted to  $(k_1 + b_1, \dots, k_n + b_n)$ . Hence the ideal  $I_V$  of  $V$  is generated by polynomials that do not depend on linear cumulants  $k_1, \dots, k_n$ . By fixing  $b_i = 0$  and moving  $a_i$ , we see that  $V$  is invariant under the torus action (10). Hence its ideal  $I_V$  is  $\mathbb{Z}^n$ -homogeneous, and this proves the only-if direction. The if-direction holds by essentially the same argument.  $\square$

The group  $SL(2)^n$  acts on the tensor space  $\mathbb{C}^{2 \times 2 \times \dots \times 2}$  and many important varieties are invariant under this action. In particular, they are invariant under  $U(2)^n$  where  $U(2)$  is the unipotent group of  $2 \times 2$ -matrices of the form

$$\begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix} \quad \text{for } \lambda \in \mathbb{C}.$$

The invariance property of Theorem 3.1 reflects precisely the  $SL(2)^n$ -invariance.

**Corollary 3.2.** *Let  $V$  be a subvariety of the affine open subset  $\{\mu_0 = 1\}$  in the projective space  $\mathbb{P}(\mathbb{C}^{2 \times 2 \times \dots \times 2})$  and let  $\bar{V}$  denote its closure in that projective*

space. If  $\bar{V}$  is invariant under the action of  $SL(2)^n$  then the ideal  $I_V$  that defines  $V$  is generated by  $\mathbb{Z}^n$ -homogeneous polynomials in the  $k_I$  with  $|I| \geq 2$ .

*Proof.* The unipotent group  $U(2)^n$  acts on the moment generating function via  $M(x) \mapsto M(x) \prod_{i=1}^n (1 + \lambda_i x_i)$ . Modulo the ideal  $\langle x_1^2, \dots, x_n^2 \rangle$  we have

$$M(x) \prod_{i=1}^n (1 + \lambda_i x_i) = M(x) \exp\left(\sum_{i=1}^n \lambda_i x_i\right) = \exp\left(K(x) + \sum_{i=1}^n \lambda_i x_i\right).$$

This means that  $U(2)^n$  acts on the space of cumulants by shifting the first order cumulants. We conclude that the prime ideal of  $V$  is generated by polynomials in the cumulants  $k_I$  with  $|I| \geq 2$ . Since  $V$  is also invariant under tuples of  $2 \times 2$ -diagonal matrices in  $SL(2)^n$ , these ideal generators can be chosen to be  $\mathbb{Z}^n$ -homogeneous.  $\square$

Hyperdeterminants and their ideals in Section 2 are  $SL(2)^n$ -invariant and hence expressible by  $\mathbb{Z}^n$ -homogeneous polynomials in higher cumulants.

**Example 3.3.** The converse does not hold in Corollary 3.2. Fix  $n = 3$ , let  $\rho \in \mathbb{C} \setminus \{4\}$ , and consider the hypersurface in  $\{\mu_\emptyset = 1\} \subset \mathbb{P}(\mathbb{C}^{2 \times 2 \times 2})$  defined by

$$k_{123}^2 + \rho \cdot k_{12} k_{13} k_{23} = 0.$$

This equation has degree six when written in the (homogenized) moments:

$$\text{Det}(P) m_\emptyset^2 + (\rho - 4)(m_\emptyset m_{12} - m_1 m_2)(m_\emptyset m_{13} - m_1 m_3)(m_\emptyset m_{23} - m_2 m_3)$$

This defines a sextic hypersurface in  $\mathbb{P}(\mathbb{C}^{2 \times 2 \times 2})$  that is  $U(2)^3$ -invariant but *not*  $SL(2)^3$ -invariant. The formula in probabilities is even less invariant:

$$\begin{aligned} & \text{Det}(P) \cdot (p_\emptyset + p_1 + p_2 + p_3 + p_{12} + p_{13} + p_{23} + p_{123})^2 \\ & + (\rho - 4)(p_\emptyset p_{23} + p_\emptyset p_{123} + p_1 p_{23} + p_1 p_{123} - p_2 p_3 - p_2 p_{13} - p_3 p_{12} - p_{12} p_{13}) \\ & \cdot (p_\emptyset p_{13} + p_\emptyset p_{123} - p_1 p_3 - p_1 p_{23} + p_2 p_{13} + p_2 p_{123} - p_3 p_{12} - p_{12} p_{23}) \\ & \cdot (p_\emptyset p_{12} + p_\emptyset p_{123} - p_1 p_2 - p_1 p_{23} - p_2 p_{13} + p_3 p_{12} + p_3 p_{123} - p_{13} p_{23}) \end{aligned}$$

Of course, for  $\rho = 4$ , this is the hyperdeterminantal quartic  $\{\text{Det}(P) = 0\}$ .  $\square$

The most basic statistical model for  $n$  binary random variables  $X_i$  is the model of *complete independence*, denoted  $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp \dots \perp\!\!\!\perp X_n$ , which is the Segre variety  $(\mathbb{P}^1)^n \subset \mathbb{P}^{2^n - 1}$ . In terms of moments, this is parametrized by  $M(x) = \prod_{i=1}^n (1 + \mu_i x_i)$ . In terms of cumulants, we obtain  $K(x) = \sum_{i=1}^n \log(1 + \mu_i x_i) = \sum_{i=1}^n k_i x_i$ . In probability coordinates, the Segre variety is defined by certain  $2 \times 2$ -determinants  $p_I p_J - p_K p_L$  but we see that this simplifies when we use cumulants as coordinates:

*Remark 3.4.* The Segre variety is defined by  $k_I = 0$  for all  $|I| \geq 2$ .

The Segre variety is the intersection of the independence models  $A \perp\!\!\!\perp B$  where  $A|B$  runs over all partitions of the set  $[n]$ . The equations for  $A \perp\!\!\!\perp B$  are  $k_I = 0$  for all  $I$  with  $A \cap I \neq \emptyset$  and  $B \cap I \neq \emptyset$ . The model  $A \perp\!\!\!\perp B$  also makes sense when  $A \cup B$  is a proper subset of  $[n]$ , with equations as follows.

**Proposition 3.5.** *If  $A$  and  $B$  are disjoint subsets of  $[n]$  then the independence model  $A \perp\!\!\!\perp B$  is defined by  $k_I = 0$  where  $I \subseteq A \cup B$ ,  $A \cap I \neq \emptyset$  and  $B \cap I \neq \emptyset$ .*

*Proof.* The independence model  $A \perp\!\!\!\perp B$  has the moment parametrization

$$M(x) = M_1(x_i : i \in A) \cdot M_2(x_j : j \in B) + \sum_{l \in [n] \setminus (A \cup B)} x_l \cdot N_l(x).$$

By taking the logarithm, we find that

$$K(x) = \log(M_1(x_i : i \in A)) + \log(M_2(x_j : j \in B)) \bmod \langle x_l : l \in [n] \setminus (A \cup B) \rangle.$$

This form is equivalent to the asserted vanishing condition on cumulants.  $\square$

The symmetry group of the  $n$ -cube is the semidirect product of the symmetric group  $S_n$ , which permutes  $[n]$ , and the abelian group  $\mathbb{Z}_2^n$ , which swaps 0s and 1s. This gives rise to an action on  $\mathbb{C}^{2 \times 2 \times \cdots \times 2}$ . We identify elements  $\rho \in \mathbb{Z}_2^n$  with subsets  $J \subseteq [n]$ . The action on coordinates  $p_I$  is as follows: for  $\sigma \in S_n$  we have  $\sigma(p_I) = p_{\sigma(I)}$ , and for  $J \subseteq [n]$  we have  $\rho_J(p_I) = p_{I \Delta J}$ , where  $I \Delta J = (I \setminus J) \cup (J \setminus I)$ . This being an action ensures

$$\rho_J(p_I) = \rho_{j_r} \circ \cdots \circ \rho_{j_1}(p_I) \quad \text{for all } J = \{j_1, \dots, j_r\} \subseteq [n]. \quad (11)$$

We extend this action from probability coordinates to any of their polynomials in a natural way. In this way we extend this action to cumulant coordinates. This action is simple for permutations  $\sigma \in S_n$ : we have  $\sigma(k_I) = k_{\sigma(I)}$ . The action of the group  $\mathbb{Z}_2^n$  is more subtle, and it can be characterized by the following corollary. That result will help us in Section 7 to get a more compact semialgebraic description of the space of cumulants, by taking advantage of the symmetries in our problem.

**Corollary 3.6.** *Consider the cumulants  $k_I$  as polynomials in probabilities  $p_I$ , via (1) and (5). For  $I, J \subseteq [n]$  with  $|I| \geq 2$ , the element  $\rho_J \in \mathbb{Z}_2^n$  satisfies*

$$\rho_J(k_I) = \begin{cases} -k_I & \text{if } |J \cap I| \text{ is odd,} \\ k_I & \text{otherwise.} \end{cases} \quad (12)$$

Furthermore, for each  $i = 1, \dots, n$ , we have

$$\rho_J(k_i) = \begin{cases} 1 - k_i & \text{if } i \in J, \\ k_i & \text{otherwise.} \end{cases}$$

*Proof.* By (11) it suffices to show  $\rho_i(k_I) = -k_I$  if  $i \in I$  and  $\rho_i(k_I) = k_I$  if  $i \notin I$ . Formula (12) follows from (10) by taking  $a_i = 0$  and  $b_i = 1$ , i.e. we swap the states of the  $i$ th variable  $X_i$ , and similarly for first-order cumulants.  $\square$

## 4. Tangents and secants of the Segre variety

In Remark 3.4 we saw that the Segre variety  $(\mathbb{P}^1)^n \subset \mathbb{P}(\mathbb{C}^{2 \times \cdots \times 2})$  collapses to a single point in the space  $\mathbb{C}^{2^n - n - 1}$  of higher cumulants. This raises the question what the representation in the  $k_I$  with  $|I| \geq 2$  looks like for varieties



naturally associated to  $(\mathbb{P}^1)^n$ , such as its secant and tangential varieties. We here examine the first tangential variety and the first secant variety:

$$\begin{aligned}\mathrm{Tan}((\mathbb{P}^1)^n) &= \text{closure of } \{x \in \mathbb{P}^{2^n-1} \mid x \text{ lies on a line tangent to } (\mathbb{P}^1)^n\}, \\ \mathrm{Sec}((\mathbb{P}^1)^n) &= \text{closure of } \{x \in \mathbb{P}^{2^n-1} \mid x \text{ lies on a secant line of } (\mathbb{P}^1)^n\}.\end{aligned}$$

Our next result reveals that the tangential variety is toric in the cumulants.

**Theorem 4.1.** *The image of the tangential variety  $\mathrm{Tan}((\mathbb{P}^1)^n)$  in the space of higher cumulants  $\mathbb{C}^{2^n-n-1}$  is isomorphic to the  $n$ -dimensional affine toric variety parametrized by all square-free monomials of degree  $\geq 2$ .*

*Proof.* In the tensor notation of [13],  $\mathrm{Tan}((\mathbb{P}^1)^n)$  has the parametrization

$$M = \frac{1}{n} \sum_{i=1}^n a^{(1)} \otimes \cdots \otimes a^{(i-1)} \otimes b^{(i)} \otimes a^{(i+1)} \otimes \cdots \otimes a^{(n)},$$

where  $a^{(i)} = (1, a_i)$  and  $b^{(i)} = (1, b_i)$  are vectors representing points in the distinguished affine open subset of  $\mathbb{P}^1$ . The formula above translates into the following parametrization of moment generating functions:

$$M(x) = \prod_{j=1}^n (1 + a_j x_j) \cdot \left( \sum_{i=1}^n \frac{1 + b_i x_i}{1 + a_i x_i} \right). \quad (13)$$

We compute the logarithm of the series  $M(x)$  modulo  $\langle x_1^2, x_2^2, \dots, x_n^2 \rangle$ . Disregarding  $\mathbb{R}$ -linear combinations of  $x_1, \dots, x_n$ , and setting  $s_i = (a_i - b_i)/n$ ,

$$K(x) = \log \left( \sum_{i=1}^n \frac{1 + b_i x_i}{1 + a_i x_i} \right) = \sum_{|I| \geq 2} (-1)^{|I|-1} (|I|-1)! \prod_{i \in I} s_i x_i, \quad (14)$$

The identity on the right can be proved directly by manipulating generating functions. An alternative and more detailed proof will be given in Example 5.2. We now conclude that

$$k_I = (-1)^{|I|-1} (|I|-1)! \cdot \prod_{i \in I} s_i \quad \text{for } |I| \geq 2. \quad (15)$$

This monomial parametrization shows that  $\mathrm{Tan}((\mathbb{P}^1)^n)$  is toric in cumulants. It is isomorphic to the toric variety with parametrization  $k_I \mapsto \prod_{i \in I} s_i$ .  $\square$

We easily find the cumulant ideal of  $\mathrm{Tan}((\mathbb{P}^1)^n)$ , by computing a Markov basis for the toric ideal of relations among squarefree polynomials of degree  $\geq 2$ . We then rescale to adjust to the signs and factorials appearing in (15).

**Example 4.2.** Let  $n = 5$ . Then the toric ideal  $\mathrm{Tan}((\mathbb{P}^1)^n)$  is minimally generated by 120 binomials in the 26 cumulant coordinates. Among these generators, 75 are quadrics and 45 are cubics. The quadrics include binomials such as  $k_{12}k_{34} - k_{14}k_{23}$ ,  $k_{123}k_{45} - k_{12}k_{345}$ ,  $k_{123}k_{345} - k_{135}k_{234}$ ,  $k_{1234} + 6k_{14}k_{23}$ , and  $k_{12345} + 12k_{12}k_{345}$ . The cubics include binomials such as  $k_{123}^2 + 4k_{12}k_{13}k_{23}$  and  $k_{123}k_{124} + 4k_{12}k_{14}k_{23}$ .  $\square$

We now come to the *secant variety*  $\text{Sec}((\mathbb{P}^1)^n)$ . This is not a toric variety in cumulants. For example, for  $n = 4$  it has the following parametrization:

$$\begin{aligned} M &= (1-t)A \otimes B \otimes C \otimes D + tE \otimes F \otimes G \otimes H, \\ M(x) &= (1-t)(1+ax_1)(1+bx_2)(1+cx_3)(1+dx_4) \\ &\quad + t(1+ex_1)(1+fx_2)(1+gx_3)(1+hx_4). \end{aligned}$$

Here  $A = (1, a), \dots, H = (1, h)$ , and  $t$  is a complex mixing parameter.

The image of  $\text{Sec}((\mathbb{P}^1)^n)$  in the 11-dimensional space of higher cumulants is a 5-dimensional affine variety that is not toric. Its ideal is generated by 16 polynomials in  $k_{12}, k_{13}, \dots, k_{1234}$ . These are the ten binomial quadrics

$$\begin{aligned} k_{12}k_{34} - k_{14}k_{23}, k_{13}k_{24} - k_{14}k_{23}, k_{12}k_{134} - k_{14}k_{123}, k_{13}k_{124} - k_{14}k_{123}, \\ k_{12}k_{234} - k_{24}k_{123}, k_{23}k_{124} - k_{24}k_{123}, k_{13}k_{234} - k_{34}k_{123}, \\ k_{23}k_{134} - k_{34}k_{123}, k_{14}k_{234} - k_{34}k_{124}, k_{24}k_{134} - k_{34}k_{124}, \end{aligned} \quad (16)$$

and the six non-binomial cubics

$$\begin{aligned} k_{12}L - (k_{123}k_{124} + 4k_{12}k_{14}k_{23}), k_{13}L - (k_{123}k_{134} + 4k_{13}k_{14}k_{23}), \\ k_{14}L - (k_{124}k_{134} + 4k_{14}k_{14}k_{23}), k_{23}L - (k_{123}k_{234} + 4k_{23}k_{14}k_{23}), \\ k_{24}L - (k_{124}k_{234} + 4k_{24}k_{14}k_{23}), k_{34}L - (k_{134}k_{234} + 4k_{34}k_{14}k_{23}). \end{aligned} \quad (17)$$

Here  $L = k_{1234} + 6k_{14}k_{23}$  is one of the toric relations on (15). Indeed, the tangential variety  $\text{Tan}((\mathbb{P}^1)^4)$  is a hypersurface in the secant variety  $\text{Sec}((\mathbb{P}^1)^4)$ . Its toric ideal in cumulants has 21 minimal generators, namely, the ten quadrics in (16), the quadric  $L$ , the six parenthesized cubics in (17), and the four hyperdeterminants

$$k_{234}^2 + 4k_{23}k_{24}k_{34}, k_{134}^2 + 4k_{13}k_{14}k_{34}, k_{124}^2 + 4k_{12}k_{14}k_{24}, k_{123}^2 + 4k_{12}k_{13}k_{23}.$$

These various equations in cumulants can now be translated back into probability coordinates, using the substitutions (1) and (5). After homogenizing and saturating with  $\mu_0$ , we recover the 32-dimensional space of  $3 \times 3$ -minors of flattenings for  $\text{Sec}((\mathbb{P}^1)^4)$ , as in [17], and the 53 ideal generators for  $\text{Tan}((\mathbb{P}^1)^4)$ , namely, the 32 cubics, the 20 hyperdeterminantal quartics, and the special quadric as in [13, §3.2].

For any  $n \geq 4$ , the secant variety  $\text{Sec}((\mathbb{P}^1)^n)$  is a curve over the toric variety  $\text{Tan}((\mathbb{P}^1)^n)$ . Using cumulant coordinates, it has the parametrization

$$k_I = \kappa_{|I|}(t) \cdot \prod_{i \in I} b_i, \quad (18)$$

where  $b_i$  are complex parameters and  $\kappa_\nu(t)$  is a certain univariate polynomial of degree  $\nu$ ; see (23). For example,

$$\begin{aligned} \kappa_2(t) &= -t^2 + t \\ \kappa_3(t) &= 2t^3 - 3t^2 + t, \\ \kappa_4(t) &= -6t^4 + 12t^3 - 7t^2 + t. \end{aligned} \quad (19)$$

The leading coefficient of  $\kappa_\nu(t)$  equals  $(-1)^{\nu-1}(\nu-1)!$  in the parametrization (15) of the tangential variety. Using (19), we can now recover the equations (16) and (17) of the secant variety by implicitizing (18) for  $n = 4$ . The derivation of (18) and the polynomials  $\kappa_\nu(t)$  will be explained in Example 5.3.

## 5. Hidden subset models

We now introduce a highly overparametrized algebraic statistical model for a vector  $X$  of  $n$  binary random variables. It is called the *complete hidden subset model*. Here is a generative description of this model. A subset  $I$  of  $[n]$  (or alternatively a binary vector) is to be chosen at random. For each element  $i \in [n]$  we need to decide whether  $i$  is in  $I$  or not. This is done as follows. First, a hidden subset  $J$  is chosen with some probability  $t_J$ . Then we select  $i$  for  $I$  with probability  $a_i^{(0)}$  if  $i \notin J$ , and we select  $i$  for  $I$  with probability  $a_i^{(1)}$  if  $i \in J$ . The conditional probabilities  $a_i^{(0)} = \text{Prob}(i \in I \mid i \notin J)$  and  $a_i^{(1)} = \text{Prob}(i \in I \mid i \in J)$  are unrelated parameters that govern this process.

The distributions in this model are parametrized as follows:

$$p_I = \sum_{J \subseteq [n]} t_J \prod_{i \in I^c \cap J^c} (1 - a_i^{(0)}) \prod_{i \in I^c \cap J} (1 - a_i^{(1)}) \prod_{i \in I \cap J^c} a_i^{(0)} \prod_{i \in I \cap J} a_i^{(1)},$$

where  $I^c$  denotes the complement of  $I \subseteq [n]$ . The corresponding moment generating function has the parametrization

$$M(x) = \sum_{J \subseteq [n]} t_J \cdot \prod_{i \in J^c} (1 + a_i^{(0)} x_i) \prod_{i \in J} (1 + a_i^{(1)} x_i). \quad (20)$$

The model has two parameters  $a_j^{(0)}$  and  $a_j^{(1)}$ , with values between 0 and 1, for each  $j \in [n]$ . Further, it has  $2^n$  mixing parameters  $t_I$ , one for each subset  $I \subseteq [n]$ . These parameters are non-negative and they sum to 1, so the tables  $T = (t_I)_{I \subseteq [n]}$  is also a distribution. We write  $k_I^{(t)}$  for the cumulants obtained from the table  $T$ .

Our main result in this section is the following intriguing theorem.

**Theorem 5.1.** *The complete hidden subset model is parametrized in terms of cumulants by  $k_i = a_i^{(0)} + b_i \cdot k_i^{(t)}$ , where  $b_i = a_i^{(1)} - a_i^{(0)}$  for  $i = 1, \dots, n$ , and*

$$k_I = k_I^{(t)} \cdot \prod_{i \in I} b_i \quad \text{for } |I| \geq 2. \quad (21)$$

*Proof.* We introduce a homogeneous probability generating function as

$$P_{\text{hom}}(y^{(0)}, y^{(1)}) = \sum_{J \subseteq [n]} p_J \prod_{i \in J^c} y_i^{(0)} \prod_{i \in J} y_i^{(1)},$$

so that  $P(x) = P_{\text{hom}}(\mathbf{1}, x)$ . Then the moment generating of  $X$  in (20) can be dually treated as the homogeneous version of the probability generating function of  $Y$ . Namely, for fixed  $a_i^{(0)}$  and  $a_i^{(1)}$ , we write (20) as  $M(x) = P_{\text{hom}}^{(t)}(y^{(0)}, y^{(1)})$ , where  $y_i^{(0)} = 1 + a_i^{(0)} x_i$  and  $y_i^{(1)} = 1 + a_i^{(1)} x_i$ . From the homogeneous generating function  $P_{\text{hom}}^{(t)}$  we can obtain the homogeneous moment generating function  $M_{\text{hom}}^{(t)}$  similarly as in the first equation in (2). Thus setting  $z_i = y_i^{(1)} - y_i^{(0)}$ , we have

$$P_{\text{hom}}^{(t)}(y^{(0)}, y^{(1)}) = P_{\text{hom}}^{(t)}(y^{(0)}, z + y^{(0)}) = M_{\text{hom}}^{(t)}(y^{(0)}, z). \quad (22)$$

From this we find that  $M(x) = P_{\text{hom}}^{(t)}(y^{(0)}, y^{(1)})$  is equal to

$$M_{\text{hom}}^{(t)}(y^{(0)}, z) = \sum_{J \subseteq [n]} \mu_J^{(t)} \prod_{i \in J} z_i \prod_{i \in J^c} y_i^{(0)}.$$

Since  $z_i = y_i^{(1)} - y_i^{(0)} = b_i x_i$  and  $y_i^{(0)} = 1 + a_i^{(0)} x_i$ , then  $M_{\text{hom}}^{(t)}(y^{(0)}, z)$  can be rewritten as a function of  $x_1, \dots, x_n$  only:

$$\begin{aligned} M_{\text{hom}}^{(t)}(y^{(0)}, z) &= \sum_{J \subseteq [n]} \mu_J^{(t)} \prod_{i \in J} b_i x_i \prod_{i \in J^c} (1 + a_i^{(0)} x_i) = \\ &= M^{(t)}(b_1 x_1, \dots, b_n x_n) \prod_{i=1}^n (1 + a_i^{(0)} x_i). \end{aligned}$$

The last equality follows holds modulo the ideal  $\langle x_1^2, \dots, x_n^2 \rangle$ . This implies  $K(x) = K^{(t)}(b_1 x_1, \dots, b_n x_n) + \sum_{i=1}^n a_i^{(0)} x_i$  and hence  $k_i = a_i^{(0)} + b_i k_i^{(t)}$  for  $i = 1, \dots, n$ , and  $k_I = k_I^{(t)} \prod_{i \in I} b_i$  for every  $I \subseteq [n]$  with  $|I| \geq 2$ .  $\square$

A *hidden subset model* is any submodel obtained from (20) by setting some of the mixing parameters  $t_I$  to zero. Thus a hidden subset model for  $n$  binary variables is specified by a collection  $\{I_1, \dots, I_k\}$  of subsets of  $[n]$ . These subsets indicate those mixing parameters  $t_{I_1}, \dots, t_{I_k}$  that are not zero. We next show that the two varieties in Section 4 arise as special cases of this.

**Example 5.2.** The hidden subset model  $\{\{1\}, \dots, \{n\}\}$  is given by

$$M(x) = \prod_{j=1}^n (1 + a_j^{(0)} x_j) \cdot \left( \sum_{i=1}^n t_i \frac{1 + a_i^{(1)} x_i}{1 + a_i^{(0)} x_i} \right).$$

This equals the tangential variety  $\text{Tan}((\mathbb{P}^1)^n)$  as in (13) but now with toric parameters  $s_i = t_i(a_i^{(1)} - a_i^{(0)})$ . We compute the cumulants  $k_I^{(t)}$  of the mixing distribution  $T$  using (5). The moments of  $T$  satisfy  $\mu_i^{(t)} = t_i$  and  $\mu_I^{(t)} = 0$  for  $|I| \geq 2$ . This means that the sum in (5) has only one non-zero term, the one corresponding to the partition  $\pi$  of  $I$  into singleton blocks. Now, (5) reads

$$k_I^{(t)} = (-1)^{|I|-1} (|I| - 1)! \prod_{i \in I} t_i$$

We have shown that the formula (21) specializes to (15) for this model.  $\square$

**Example 5.3.** The hidden subset model  $\{\emptyset, [n]\}$  is the mixture of two independent random variables, so it coincides with the secant variety  $\text{Sec}((\mathbb{P}^1)^n)$ . The mixing distribution  $T$  has one free mixing parameter  $t$ , where  $t_{1\dots n} = t$  and  $t_\emptyset = 1 - t$ . The moments of  $T$  are  $\mu_\emptyset = 1$  and  $\mu_B^{(t)} = t$  for  $|B| \geq 1$ . The formula (5) implies

$$k_I^{(t)} = \sum_{i=1}^{|I|} (-1)^{i-1} (i-1)! \cdot \beta_{i,|I|} \cdot t^i$$

where  $\beta_{i,|I|}$  is the number of set partitions of  $I$  into  $i$  blocks. This univariate polynomial depends only on the cardinality  $\nu = |I|$ . We can also write it as

$$\kappa_\nu(t) = \sum_{i=1}^{\nu} (-1)^{i-1} \cdot \gamma_{i,\nu} \cdot t^i \quad (23)$$

where  $\gamma_{i,I}$  is the number of cyclically ordered set partitions of a  $\nu$ -set into  $i$  blocks. Such partitions are known as *necklaces* in enumerative combinatorics.  $\square$

**Example 5.4 (Binary Hidden Markov Model).** The complete hidden subset model includes all *hidden Markov models* (HMM) where both hidden and observed states are binary. These models are widely used in computational biology [15, §1.4.3 and §11]. We can treat the mixing variable with distribution  $T$  as a hidden binary process  $Y = (Y_1, \dots, Y_n)$ . The parameters  $a_i^{(0)}$  and  $a_i^{(1)}$  determine the conditional distribution of  $X_i$  given the hidden process, and this observed distribution depends on  $Y$  only through the value of  $Y_i$ . In this context the parameters  $b_i = a_i^{(1)} - a_i^{(0)}$  are the linear regression coefficients of  $X_i$  with respect to  $Y_i$ . For an HMM, the hidden distribution  $T$  follows, in addition, a homogeneous Markov chain [15, §1.4.2]. Thus, if  $k_I^{(t)}$  are the cumulants of the homogeneous Markov chain, then (21) gives a parametrization of the binary HMM. It would be interesting to revisit the recent work of Schönhuth [22] from this perspective. We expect his prime ideals  $I_{3,n}$  in [22, §7.3] to have a nice representation in terms of cumulants.  $\square$

The set of all hidden subset models, for fixed  $n$ , forms a poset whose elements  $\mathcal{M}_A$  are indexed by the  $2^{2^n}$  subsets  $A$  of  $2^{[n]}$ . The model  $\mathcal{M}_A$  is obtained from the complete hidden subset model by setting  $t_I = 0$  for all  $I \subseteq [n]$  not in  $A$ . Of course, different labels  $A$  and  $B$  can lead to isomorphic hidden subset models  $\mathcal{M}_A$  and  $\mathcal{M}_B$ . Clearly, this happens if  $B$  is obtained from  $A$  by a permutation of  $[n]$ . But, in fact, the full symmetry cube of the  $n$ -cube acts on the hidden subset models:

**Proposition 5.5.** *Let  $A, B \subseteq 2^{[n]}$  and assume that  $B$  is equal to  $J\Delta A = \{I\Delta\alpha : \alpha \in A\}$  for some subset  $J \subseteq [n]$ . Then  $\mathcal{M}_A$  is isomorphic to  $\mathcal{M}_B$ .*

*Proof.* By Theorem 5.1, the two models are parametrized by  $k_I = k_I^{(t)} \prod_{i \in I} b_i$ . The cumulants  $k_I^{(t)}$  depend on  $A$  and  $B$ . By Corollary 3.6, if  $B = J\Delta A$  for some  $J$ , then the respective cumulants  $k_I^{(t)}$  for  $A$  and  $B$  agree up to sign.  $\square$

Each hidden subset model  $\mathcal{M}_A$  can be identified with a 0/1-polytope  $P_A$ . By Proposition 5.5, if  $P_A$  and  $P_B$  are 0/1 equivalent then  $\mathcal{M}_A$  and  $\mathcal{M}_B$  are isomorphic. We say that the model  $\mathcal{M}_A$  is *non-degenerate* if the polytope  $P_A$  is not contained in any hyperplane  $x_i = 0$  or  $x_i = 1$  for  $i = 1, \dots, n$ . If this happens then the random variable  $X_i$  is independent of all other variables. Geometrically this means that the variety of  $\mathcal{M}_A$  decomposes as a product of  $\mathbb{P}^1$  and a smaller hidden subset model.

If  $n = 2$  then, up to the symmetry of the 2-cube, there are precisely three distinct hidden subset models which are non-degenerate:  $\{\emptyset, 1, 2, 12\}$ ,  $\{\emptyset, 1, 2\}$ ,  $\{\emptyset, 12\}$ . Their models  $\mathcal{M}_A$  all parametrize the full tetrahedron  $\Delta_3$  of distributions on  $2^{[2]}$  or, in algebraic terms, the whole projective space  $\mathbb{P}^3$ .

If  $n = 3$  then, up to symmetry of the 3-cube, there are precisely 19 collections  $A$  of subsets of  $\{1, 2, 3\}$  with  $2 \leq |A| \leq 7$ . Thirteen of these 19

models  $\mathcal{M}_A$  have codimension 0, that is, they are full-dimensional in the simplex  $\Delta_7$  of probability distributions on  $2^{[3]}$ . One of these sets is  $A = \{\emptyset, 123\}$  which represents  $\mathcal{M}_A = \text{Sec}((\mathbb{P}^1)^3)$  and hence fills  $\mathbb{P}^7$ . The remaining six of the 19 models  $\mathcal{M}_A$  represent three distinct varieties. The first of them is the hyperdeterminantal hypersurface  $\text{Tan}((\mathbb{P}^1)^3)$ . The other two varieties are a line and a point in cumulant space:

hidden subset model	variety	codimension
$\{\emptyset, 1, 2, 3\}$ or $\{\emptyset, 12, 13\}$	$k_{123}^2 + 4k_{12}k_{13}k_{23} = 0$	1
$\{1, 2\}$ or $\{\emptyset, 1, 2\}$ or $\{\emptyset, 1, 2, 12\}$	$(k_{12}, k_{13}, k_{23}, k_{123}) = (0, 0, 0, *)$	3
$\{\emptyset, 1\}$	$(k_{12}, k_{13}, k_{23}, k_{123}) = (0, 0, 0, 0)$	4

The first row corresponds to non-degenerate models  $\mathcal{M}_A$  that do not fill  $\Delta_7$ . The situation becomes more interesting for  $n \geq 4$  when we get a vast range of new models. Some of these will be discussed and catalogued in Section 6.

## 6. Context-specific independence

This section concerns a class of statistical models that has proved to be useful in machine learning and computational biology [6], namely, the context-specific independence (CSI) for binary random variables. It has been observed in [13, §6.3] that both the tangential variety and the secant variety of the Segre variety are CSI models. Examples 5.2 and 5.3 expressed these as hidden subset models. We here generalize this relationship by identifying the class of binary CSI models with a natural class of hidden subset models.

The formal specification of a CSI model is as follows. Fix a multiset of  $n$  partitions  $\{\pi_1, \pi_2, \dots, \pi_n\}$  of the set  $[m] = \{1, \dots, m\}$ . The model is

$$M(x) = \sum_{j=1}^m t_j (1 + a_{\pi_1(j)} x_1) (1 + b_{\pi_2(j)} x_2) \cdots (1 + c_{\pi_n(j)} x_n). \quad (24)$$

Here  $\pi_j(k)$  is the block of the  $j$ -th partition  $\pi_j$  that contains the class  $k$ , and  $t_1, \dots, t_m$  are mixing parameters for the classes. These satisfy  $t_1 + \dots + t_m = 1$ .

If each  $\pi_j$  is the partition into singletons, then we can write  $\pi_i(j) = j$  in (24) and the CSI model is the  $m$ -th secant variety of  $(\mathbb{P}^1)^n$  in  $\mathbb{P}^{2^n-1}$ . This is known in statistics as mixture of  $n$  independent binary vectors or as the *naive Bayes model*. Hence every CSI model with  $m$  hidden classes is a submodel of the naive Bayes model.

The CSI model has  $\sum_{i=1}^n |\pi_i| + m - 1$  parameters and the dimension of the ambient space is  $2^n - 1$ . It is usually not identifiable, meaning its dimension is smaller than the number of parameters. However, identifiability does hold for  $m = 2$ . Here the CSI model is the product of the Segre variety  $(\mathbb{P}^1)^{n-k}$  and the first secant variety of  $(\mathbb{P}^1)^k$ , where  $k = \#\{j \in [n] : \pi_j = 1|2\}$ . This is the graphical model represented by a directed star tree with a hidden binary variable and  $k$  leaves together with  $n - k$  isolated nodes.

From statistical point of view it is sensible to assume the following:

- (A1) All partitions in the model specification have at least two blocks.

- (A2) There is no pair of elements  $\{i, j\}$  such that for every partition in the model specification both  $i$  and  $j$  are in the same block of this partition.

If (A1) is violated then one random variable is independent of all others. Taking an appropriate margin, we can constrain our analysis to the remaining variables. If (A2) does not hold then the classes  $i$  and  $j$  can be joined to form a single class without changing the model. If  $m = n = 3$  then up to symmetry we have three CSI models satisfying (A1) and (A2). The first case is  $\pi_1 = 1|23$   $\pi_2 = 2|13$ , and  $\pi_3 = 3|12$ . This is precisely our hyperdeterminantal hypersurface  $\text{Tan}((\mathbb{P}^1)^3) = V(k_{123}^2 + 4k_{12}k_{13}k_{23})$ . The other two CSI models represent all distributions on  $2^{[3]}$ :

$$(\pi_1 = 1|23, \pi_2 = 12|3, \pi_3 = 1|2|3) \quad \text{or} \quad (\pi_1 = 1|23, \pi_2 = \pi_3 = 1|2|3).$$

In the remainder of this section we study a special class of CSI models. Namely, we shall require that each partition  $\pi_i$  has precisely two blocks. We call these models the *CSI split models*. Thus a CSI split model is represented by a collection  $\{\pi_1, \pi_2, \dots, \pi_n\}$  of splits of the set  $[m]$  of hidden states. The following result identifies these models with the models in Section 5.

**Proposition 6.1.** *The CSI split models are precisely the hidden subset models.*

*Proof.* Let  $\mathcal{M}_A$  be the hidden subset model defined by  $A = \{J_1, \dots, J_m\} \subseteq 2^{[n]}$ . This is written as a CSI split model with  $m$  hidden classes by taking the  $n$  partitions  $\pi_1, \dots, \pi_n$  of  $[m]$  to be  $\pi_i = I|I^c$ , where  $\ell \in I$  whenever  $i \in J_\ell$ . Conversely, suppose we are given a CSI split model  $\{\pi_1, \dots, \pi_n\}$ . Then we regard  $\pi_i$  as an ordered partition, and we recover the  $m$  subsets in  $A$  by taking  $J_\ell$  to be the set of all  $i \in [n]$  such that  $\ell$  is in the first part of  $\pi_i$ . These transformations lead to identical parametrizations, and hence the corresponding models in  $\Delta_{2^n-1}$  coincide.  $\square$

We classified all hidden subset models and hence all CSI split models for  $n = 3$  in the end of the previous section. The next case  $n = 4$  is much more interesting, as it offers a considerably wider range of possibilities. The classification for  $n = 4$  will occupy us in the rest of this section.

**Example 6.2 (The hyperdeterminant as CSI model).** Let  $n = 4$ ,  $m = 7$ , and consider the hidden subset model  $\mathcal{M}_A$  where  $A = \{\emptyset, 12, 13, 14, 23, 24, 34\}$ . The corresponding CSI split model is  $\{1234|567, 1256|347, 1357|246, 1467|235\}$ . In algebraic geometry, the model  $\mathcal{M}_A$  corresponds to the *second osculating variety* of the Segre variety  $(\mathbb{P}^1)^4$ . This is a hypersurface of degree 24 in  $\mathbb{P}^{15}$ , namely, it is the hypersurface defined by the  $2 \times 2 \times 2 \times 2$ -hyperdeterminant. This result was pointed out to us by Luke Oeding and Giorgio Ottaviani, and we can easily verify it by a direct computation. The fact that  $\text{codim}(\mathcal{M}_A) = 1$  is verified by computing the rank of the Jacobian of the parametrization (21) for random parameter values. The fact that  $\mathcal{M}_A$  equals  $\{\text{Det}(P) = 0\}$  is verified by plugging the parametrization (21) into the formula with 13819 monomials found in Theorem 2.1. We note that this model remains the same if we augment  $A$  to  $\{\emptyset, 1, 2, 3, 4, 12, 13, 14, 23, 24, 34\}$ .  $\square$

Hidden subset model	CSI split model	codimension	degree
$\{\emptyset, 12, 13, 14\}$	$1 234, 2 134, 3 124, 4 123$	7	20
$\{\emptyset, 12, 13, 4\}$	$14 23, 2 134, 3 124, 4 123$	6	29
$\{\emptyset, 1, 2, 34\}$	$2 134, 3 124, 4 123, 4 123$	6	29
$\{\emptyset, 1, 23, 234\}$	$2 134, 34 12, 34 12, 4 123$	6	23
$\{\emptyset, 1, 234, 1234\}$	$24 13, 34 12, 34 12, 34 123$	6	23
$\{\emptyset, 1, 2, 134\}$	$13 24, 3 124, 4 123, 4 123$	5	44
$\{\emptyset, 1, 12, 234\}$	$23 14, 34 12, 4 123, 4 123$	5	44
$\{\emptyset, 1, 123, 234\}$	$23 14, 34 12, 34 12, 4 123$	5	44
$\{\emptyset, 1, 23, 124\}$	$24 13, 34 12, 3 124, 4 123$	5	31
$\{\emptyset, 12, 134, 234\}$	$23 14, 24 13, 34 12, 34 12$	5	22
$\{\emptyset, 12, 13, 24\}$	$14 23, 13 24, 3 124, 4 123$	4	44
$\{\emptyset, 13, 23, 124\}$	$13 24, 12 34, 14 23, 4 123$	4	38
$\{\emptyset, 12, 34, 1234\}$	$24 13, 24 13, 34 12, 34 12$	4	11
$\{\emptyset, 1, 234\}$	$2 13, 3 12, 3 12, 3 12$	6	23
$\{\emptyset, 12, 134\}$	$1 23, 2 13, 3 12, 3 12$	6	29
$\{\emptyset, 12, 34\}$	$2 13, 2 13, 3 12, 3 12$	5	44
$\{\emptyset, 1234\}$	$1 2, 1 2, 1 2, 1 2$	6	23

TABLE 1. The 17 non-degenerate CSI split models on  $n=4$  binary variables with  $m \leq 4$  hidden classes, up to symmetry.

We now come to the classification of CSI split models for  $n = 4$ . Each model lives in the space  $\mathbb{C}^{11}$  with coordinates  $k_{12}, \dots, k_{34}, k_{123}, \dots, k_{234}, k_{1234}$ .

**Proposition 6.3.** *Up to symmetry, for  $n=4$ , there are 380 CSI split models satisfying (A1) and (A2). The number of models with  $m$  hidden classes is*

$$0, 1, 3, 13, 24, 47, 55, 73, 56, 50, 27, 19, 6, 4, 1, 1 \quad \text{for } m = 1, 2, \dots, 16.$$

*In Table 1 we list the codimension and degree for all 17 models for  $m \leq 4$ .*

For our classification we used the representation of each CSI split model as a hidden subset model  $\mathcal{M}_A$ , where  $A \subseteq 2^{[4]}$ , given by Proposition 6.1. A choice of  $A$  is also displayed for each model in Table 1. Note that two distinct models may define the same variety. By symmetry we can assume  $\emptyset \in A$ . We first generated the list of all non-degenerate sets  $A$  of subsets of  $\{1, 2, 3, 4\}$  containing  $\emptyset$ , we then computed orbits under the symmetry group of the 4-cube, and finally we selected one representative per orbit. To compute the codimension  $c$  of  $\mathcal{M}_A$ , we evaluated the rank of the Jacobian of the polynomial map (21) at random values of the parameters. By *degree* in Proposition 6.3 we mean the number of complex solutions on  $\mathcal{M}_A$  of a system of  $11 - c$  inhomogeneous linear equations with random coefficients in the 11 unknowns  $k_I$ . We used `Macaulay2` [7] to count the number of solutions to these equations. It would be desirable to compute the defining prime ideals for all models in Proposition 6.3, but we found this to be difficult for  $m \geq 5$ .



The 380 models represent a nice suite of test problems for *implicitization* in computer algebra. We close the section with one easy instance.

**Example 6.4.** The model  $\mathcal{M}_A$  with  $A = \{\emptyset, 12, 34, 1234\}$  has CSI representation  $\{12|34, 12|34, 13|24, 13|24\}$ . Its prime ideal in cumulant coordinates is

$$\langle k_{13}k_{24} - k_{14}k_{23}, k_{13}k_{124} - k_{14}k_{123}, k_{13}k_{234} - k_{23}k_{134}, k_{14}k_{234} - k_{24}k_{134}, \\ k_{23}k_{124} - k_{24}k_{123}, k_{23}k_{1234} - k_{234}k_{123} + 2k_{14}k_{23}^2, k_{13}k_{1234} - k_{134}k_{123} + 2k_{14}k_{13}k_{23}, \\ k_{23}k_{1234} - k_{234}k_{124} + 2k_{14}k_{24}k_{23}, \text{ and } k_{14}k_{1234} - k_{134}k_{124} + 2k_{14}^2k_{23} \rangle.$$

This CSI split model has codimension 4 and degree 11.  $\square$

## 7. Semialgebraic geometry of the space of cumulants

In the previous sections we studied binary cumulant varieties as objects of complex algebraic geometry. We examined their dimension, parameterization, and defining prime ideal, but we largely ignored the issue that parameters and probabilities are real and non-negative. In statistical applications, however, it is essential to work over the real numbers and to pay attention to the pertinent inequalities. In this section seek to address this omission by asking the following fundamental question: Which  $2 \times 2 \times \cdots \times 2$ -tables  $K = (k_I)_{I \subseteq [n]}$  with entries in the real numbers represent the cumulants of actual probability distributions  $P = (p_I)_{I \subseteq [n]}$ ?

Our object of study is the image of the polynomial map  $\Delta_{2^n-1} \rightarrow \mathbb{R}^{2^n-1}$  taking probability distributions  $P$  to their cumulants  $K$ . This image is denoted  $\mathcal{K}_n$ . We call it the *space of cumulants*. The space of cumulants  $\mathcal{K}_n$  is a semi-algebraic subset of  $\mathbb{R}^{2^n-1}$ . This means that it has a description in terms of polynomial inequalities in the  $k_I$ . We begin by offering a convenient representation of these inequalities.

**Proposition 7.1.** *The space of cumulants  $\mathcal{K}_n$  is a basic semialgebraic set in  $\mathbb{R}^{2^n-1}$ . It consists of the solutions of the polynomial inequalities*

$$\sum_{\pi \in \Pi([n])} \prod_{B \in \pi} \rho_J(k_B) \geq 0 \quad \text{for all } J \subseteq [n]. \quad (25)$$

*Proof.* The set  $\mathcal{K}_n$  being *basic semialgebraic* means that it is described by a finite conjunction of polynomial inequalities. That conjunction is (25), and we shall now prove it. The moment  $\mu_{1\dots n}$  agrees with the probability  $p_{1\dots n}$ , so it is non-negative on  $\mathcal{K}_n$ . Expressing  $\mu_{1\dots n}$  in terms of cumulants as in (6),

$$p_{1\dots n} = \sum_{\pi \in \Pi([n])} \prod_{B \in \pi} k_B \geq 0.$$

By applying the transformation  $\rho_J$  from (11) to this inequality, we obtain  $p_{J^c} = \rho_J(p_{1\dots n}) \geq 0$ . This translates into the inequality (25) in cumulants. Since the transformation  $P \mapsto K$  is invertible, we see that  $\mathcal{K}_n$  has the desired representation.  $\square$

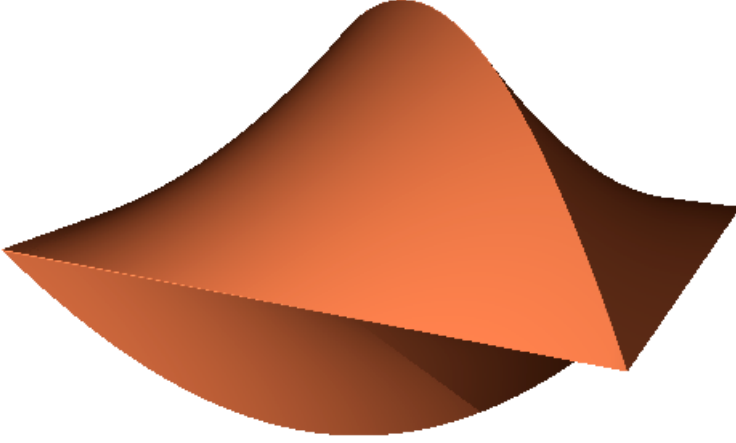


FIGURE 1. The space of cumulants  $\mathcal{K}_2$  is the solution set of (26)

**Example 7.2 (Space of cumulants for  $n = 2$ ).** The probability distributions  $P$  on the subsets of  $\{1, 2\}$  form a tetrahedron, and we map this tetrahedron into the 3-space with coordinates  $(k_1, k_2, k_{12})$ . The image of this map is the space of cumulants  $\mathcal{K}_2$ . Proposition 7.1 gives the semialgebraic representation:

Inequalities defining $\Delta_3$ $p_{12} \geq 0$ $p_1 \geq 0$ $p_2 \geq 0$ $p_\emptyset \geq 0$	Inequalities defining $\mathcal{K}_2$ $k_{12} \geq -k_1 k_2,$ $k_{12} \leq k_2(1 - k_1),$ $k_{12} \leq k_1(1 - k_2),$ $k_{12} \geq -(1 - k_1)(1 - k_2).$	(26)
---	--	------

The solution set of these four quadratic inequalities is depicted in Figure 1. In this diagram we see clearly how  $\mathcal{K}_2$  arises as a non-linear image of the tetrahedron  $\Delta_3$ . Note that the body  $\mathcal{K}_2$  is not convex. The square  $\{0 \leq k_1, k_2 \leq 1\}$  in the plane  $\{k_{12} = 0\}$  is the image of the independence model  $\{p_\emptyset p_{12} = p_1 p_2\}$ , which contains four of the six edges of  $\Delta_3$ . The other two edges of the tetrahedron  $\Delta_3$  are the quadratic curves that form the ridges at the top and the bottom of the  $\mathcal{K}_2$ .  $\square$

**Example 7.3 (The space of cumulants for  $n = 3$ ).** We now consider the simplex  $\Delta_7$  of distributions on subsets of  $\{1, 2, 3\}$ . Its image in cumulant coordinates is the 7-dimensional closed basic semialgebraic set  $\mathcal{K}_3$ . Both  $\Delta_7$  and  $\mathcal{K}_3$  are defined by the constraints that the following eight expressions should be non-negative:

$$\begin{aligned}
 p_{123} &= \mu_{123} &= k_{123} + k_{12}k_3 + k_{13}k_2 + k_{23}k_1 + k_1k_2k_3 \\
 p_{12} &= -\mu_{123} + \mu_{12} &= -k_{123} - k_{12}(k_3 - 1) - k_{13}k_2 - k_{23}k_1 - k_1(k_2 - 1)k_3 \\
 p_{13} &= -\mu_{123} + \mu_{13} &= -k_{123} - k_{12}k_3 - k_{13}(k_2 - 1) - k_{23}k_1 - (k_1 - 1)k_2k_3 \\
 p_{23} &= -\mu_{123} + \mu_{23} &= -k_{123} - k_{12}k_3 - k_{13}k_2 - k_{23}(k_1 - 1) - (k_1 - 1)k_2k_3
 \end{aligned}$$

$$\begin{aligned}
p_1 &= \mu_{123} - \mu_{12} - \mu_{13} + \mu_1 = k_{123} + k_{12}(k_3 - 1) + k_{13}(k_2 - 1) + k_{23}k_1 + k_1(k_2 - 1)(k_3 - 1) \\
p_2 &= \mu_{123} - \mu_{12} - \mu_{23} + \mu_2 = k_{123} + k_{12}(k_3 - 1) + k_{13}k_2 + k_{23}(k_1 - 1) + (k_1 - 1)k_2(k_3 - 1) \\
p_3 &= \mu_{123} - \mu_{13} - \mu_{23} + \mu_3 = k_{123} + k_{12}k_3 + k_{13}(k_2 - 1) + k_{23}(k_1 - 1) + (k_1 - 1)(k_2 - 1)k_3 \\
p_0 &= -\mu_{123} + \mu_{12} + \mu_{13} + \mu_{23} - \mu_1 - \mu_2 - \mu_3 + 1 \\
&= -k_{123} - k_{12}(k_3 - 1) - k_{13}(k_2 - 1) - k_{23}(k_1 - 1) - (k_1 - 1)(k_2 - 1)(k_3 - 1)
\end{aligned}$$

Thus the space  $\mathcal{K}_3$  is defined by eight cubic inequalities in  $\mathbb{R}^7$ .  $\square$

Equipped with the inequality description of  $\mathcal{K}_n$  we can now try to answer questions about the geometry of cumulants of probability distributions. One natural such question is to identify the smallest box containing  $\mathcal{K}_n$ . This is equivalent to find a tight upper and lower bound on the possible values of the cumulants  $k_I$ . The following problem was suggested by Gian-Carlo Rota and his collaborators in [1]:

$$\text{Maximize } |k_{12\dots n}| \text{ subject to } k \in \mathcal{K}_n. \quad (27)$$

In this problem, the absolute value sign around  $k_{12\dots n}$  can be removed because  $k \in \mathcal{K}_n$  implies  $-k \in \mathcal{K}_n$ . This is shown in [1, Proposition 3.1], and it also follows directly from the symmetries in Corollary 3.6. Let  $\mathbf{k}_n^*$  denote the optimal value of (27). Figure 1 shows that  $\mathbf{k}_2^* = 1/4$ , and one easily derives an algebraic proof from the inequalities in Example 7.2. The probability distribution  $p_0 = p_{12} = \frac{1}{2}$  attains  $\mathbf{k}_2^* = 1/4$ . It has been conjectured by Bruno, Rota and Torney [1] that the analogous distribution solves the optimization problem (27) for all even values of  $n$ :

**Conjecture 7.4.** [1, bottom of page 16] If  $n \geq 2$  is an even integer then

$$\mathbf{k}_n^* = \kappa_n\left(\frac{1}{2}\right) = (-1)^{n/2} \sum_{i=1}^n \left(-\frac{1}{2}\right)^i \cdot \gamma_{i,n}.$$

This value is attained by the probability distribution  $p_0 = p_{1\dots n} = \frac{1}{2}$ .

Here  $\kappa_n$  is the polynomial in (23). The first values of the bound  $\kappa_n(\frac{1}{2})$  are  $\frac{1}{4}, \frac{1}{8}, \frac{1}{4}, \frac{17}{16}, \frac{31}{4}$  for  $n = 2, 4, 6, 8, 10$ . It has been remarked in [1] that

$$\left|\kappa_n\left(\frac{1}{2}\right)\right| \sim 2 \frac{1}{\pi^{2n}} (2n-1)! \quad \text{for } n \gg 0.$$

If  $n \geq 3$  is an odd integer then  $\kappa_n(\frac{1}{2}) = 0$ , and no conjectured value for  $\mathbf{k}_n^*$  has been suggested in [1]. Using recent computational advances in certified polynomial optimization, we attacked the problem (27) for  $n = 3$  and  $n = 4$ , thus confirming the conjecture of Bruno, Rota and Torney in the first non-trivial case. Namely, we found that the upper bound on cumulants of probability set functions satisfies

$$\mathbf{k}_3^* = \mathbf{k}_4^* = \frac{1}{8}. \quad (28)$$

For  $n = 3$  we used the software **Bermeja** [19] to compute a *sums of squares* certificate via semidefinite programming. We are grateful for the help

provided by Philipp Rostalski. Let us now explain this certificate. We consider the following cubic polynomial in the seven moment coordinates  $\mu_I$ :

$$\frac{1}{8} - k_{123} = \frac{1}{8} - \mu_{123} + \mu_1\mu_{23} + \mu_2\mu_{12} + \mu_3\mu_{12} - 2\mu_1\mu_2\mu_3.$$

Our aim is to prove that this polynomial is non-negative on the simplex  $\Delta_7$ . We do this by rewriting the polynomial in the following special form

$$\frac{1}{8} - k_{123} = \sigma_\emptyset + \sigma_1\mu_1 + \sigma_2\mu_2 + \sigma_3\mu_3 + \sigma_{12}\mu_{12} + \sigma_{13}\mu_{13} + \sigma_{23}\mu_{23} + \sigma_{123}\mu_{123}, \quad (29)$$

where each of the eight multipliers  $\sigma_I$  is a sum of squares of linear polynomial in the moments  $\mu_J$ . Each such sum of squares corresponds to a positive semidefinite quadratic form, and it can be represented by a symmetric  $8 \times 8$ -matrix  $\Sigma_I$  as follows:

$$\sigma_I = \mu \cdot \Sigma_I \cdot \mu^T \quad \text{where} \quad \mu = (1, \mu_1, \mu_2, \mu_3, \mu_{12}, \mu_{13}, \mu_{23}, \mu_{123}). \quad (30)$$

Our certificate for  $\mathbf{k}_3^* = 1/8$  is a tuple  $(\Sigma_\emptyset, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_{12}, \Sigma_{13}, \Sigma_{23}, \Sigma_{123})$  of positive semidefinite symmetric  $8 \times 8$ -matrices such that (29) and (30) hold. Finding such a tuple of matrices is an instance of semidefinite programming.

We attempted to find a similar proof for the second identity  $\mathbf{k}_4^* = 1/8$  but the computations required turned out to be too difficult so far. The idea was to take advantage of the symmetries preserves the optimization problem (27). This is a group of order 192, and has index 2 in the symmetry group of the 4-cube. Our hope was to use the the dual moment formulation due to Riener *et al.* in [18], but this did yet terminate successfully. Instead, we verified the identify  $\mathbf{k}_4^* = 1/8$  by running numerous applications of standard implementations of numerical optimization in **R** and **Matlab**. Running these hill climbing methods from a multitude of different starting values verifies the desired result with very high confidence.

## References

- [1] W. BRUNO, G. ROTA, AND D. TORNEY: *Probability set functions*, Annals of Combinatorics, 3 (1999), pp. 13–25.
- [2] D. BRUYNNOOGHE AND H. WYNN: Differential cumulants, hierachical models and monomial ideals, [arXiv:1102.2118](https://arxiv.org/abs/1102.2118).
- [3] M. DRTON, B. STURMFELS, AND S. SULLIVANT: *Lectures on Algebraic Statistics*, Oberwolfach Seminars, Vol. 39, Birkhäuser, Basel, 2009.
- [4] M. J. DUFF, *String triality, black hole entropy, and Cayley’s hyperdeterminant*, Phys. Rev. D, 76 (2007), 025017, 4.
- [5] I.M. GEL’FAND, M.M. KAPRANOV, AND A.V. ZELEVINSKY: *Discriminants, Resultants, and Multidimensional Determinants*, Birkhäuser, 1994.
- [6] B. GEORGI AND A. SCHLIEP: *Context-specific independence mixture modeling for positional weight matrices*, Bioinformatics, 22 (2006), pp. e166–e173.
- [7] D. GRAYSON AND M. STILLMAN: *Macaulay 2* – a software system for research in algebraic geometry, Available at <http://www.math.uiuc.edu/Macaulay2/>.
- [8] A. HALD, *The early history of the cumulants and the Gram-Charlier series*, International Statistical Review, 68 (2000), pp. 137–153.

- [9] O. HOLTZ AND B. STURMFELS: *Hyperdeterminantal relations among symmetric principal minors*, Journal of Algebra, 316 (2007), pp. 634–648.
- [10] P. HUGGINS, B. STURMFELS, J. YU, AND D. YUSTER: *The hyperdeterminant and triangulations of the 4-cube*, Mathematics of Computation, 77 (2008), pp. 1653–1680.
- [11] P. MCCULLAGH: *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1987.
- [12] A. MIYAKE, *Classification of multipartite entangled states by multidimensional determinants*, Phys. Rev. A (3), 67 (2003), 012108, 10.
- [13] L. OEDING: *Set-theoretic defining equations of the tangential variety of the Segre variety*, J. Pure and Applied Algebra, 215 (2011), pp. 1516–1527.
- [14] L. OEDING: *Set-theoretic defining equations of the variety of principal minors of symmetric matrices*, Algebra and Number Theory, to appear.
- [15] L. PACTER AND B. STURMFELS: *Algebraic Statistics for Computational Biology*, Cambridge University Press, 2005.
- [16] G. PISTONE AND H.P. WYNN: *Cumulant varieties*, J. Symbolic Computation, 41 (2006), pp. 210–221.
- [17] C. RAICU: *The GSS conjecture*, [arXiv:1011.5867](https://arxiv.org/abs/1011.5867).
- [18] C. RIENER, T. THEOBALD, L. JANSSON ANDRÉN AND J.B. LASSERRE: *Exploiting symmetries in SDP-relaxations for polynomial optimization*, [arXiv:1103.0486](https://arxiv.org/abs/1103.0486).
- [19] P. ROSTALSKI: *Bermeja* – software for convex algebraic geometry, available at [math.berkeley.edu/~philipp/cagwiki](http://math.berkeley.edu/~philipp/cagwiki).
- [20] G.-C. ROTA AND S. ROMAN, *The umbral calculus*, Adv. Math, 27 (1978), pp. 95–188.
- [21] G.-C. ROTA AND J. SHEN: *On the combinatorics of cumulants*, J. Combin. Theory Ser. A, 91 (2000), pp. 283–304.
- [22] A. SCHÖNHUTH: *Complete identification of binary-valued hidden Markov processes*, [arXiv:1101.3712](https://arxiv.org/abs/1101.3712).
- [23] T.P. SPEED: *Cumulants and partition lattices*, Austral. J. Statistics, 25 (1983), pp. 378–388.
- [24] R.P. STANLEY: *Enumerative Combinatorics. Volume I*, no. 49 in Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2002.
- [25] J.G. SUMNER AND P.D. JARVIS: *Using the tangle: a consistent construction of phylogenetic distance matrices for quartets*, Mathematical Biosciences, 204 (2006), pp. 49–67.
- [26] D.C. TORNEY: *Binary cumulants*, Advances in Applied Math, 25, (2000), pp. 34–40.

Bernd Sturmfels

Department of Mathematics, University of California, Berkeley, CA 94720, USA,  
[bernd@math.berkeley.edu](mailto:bernd@math.berkeley.edu)

Piotr Zwiernik

Institute Mittag-Leffler, 18260 Djursholm, Sweden, [piotr.zwiernik@gmail.com](mailto:piotr.zwiernik@gmail.com)